Role of Statistical Learning Theory on Probabilistic Robust Controller Synthesis Problem

Titas Bera, M. Seetharama Bhat and D. Ghose

***Abstract*— In this paper we review the randomized algorithm based design philosophy for probabilistic robust controller synthesis and the role of empirical process theory on designing such algorithms. Among the existing polynomial time randomized algorithms some have over conservative sample complexity bounds and consequently there lies considerable scope to improve those bounds.**

I. INTRODUCTION

During recent years it has been reported that several problem in robustness analysis and synthesis are either NP complete or NP hard. A survey of NP hard problems in control theory can be found in the work of Blondel and Tsitkilis [1] and [2], Bernstein [3], Nemirovskii [4], Poljak and Rohn [5]. These results show that various problems in robust control are practically unsolvable if the number of variables become sufficiently large. The problem of checking robust stability and performance under structured / parametric uncertainty is proven to be NP-hard. However, parametric uncertainty can also be handled in the H_∞ framework, but then the design will be too conservative. Although, μ analysis / synthesis approach although can handle parametric uncertainty, but computational complexity of μ calculation is again NP-hard. Also, stability evaluation of interval matrices is itself a NP-hard problem. For details of computational complexity of algorithms see [6].

Recent research on computational complexity of robust control analysis and design problems indicates that these difficulties are most likely, inherent to the problem formulations rather than a lack of ingenuity. To overcome the intractability issues and conservatism, recent trends are to consider the problem of stability and performance robustness in a probabilistic framework. One can refer to [7] for details of these approaches. Among others, the statistical learning theory based algorithms, originally initiated by Vidyasagar [8] is of considerable interest. In this framework one designs a controller based on average performance requirement as the plant varies over a pre-specified family and is not based on worst case performance requirements. For such cases, the controller synthesis problem can be formulated as the minimization of expected value of the objective function. Statistical learning theory shows that whenever a property known as uniform convergence of empirical means holds, there exists an efficient randomized algorithm for an associated function minimization problem. Therefore, one can derive sample size estimates to approximate a minimum of a function with arbitrary accuracy and confidence. The approach can tackle a wide range of cost functions and is not necessarily confined to only convex functions. Some specific problems that fall within this framework include robust stabilization of a family of plants and minimization of weighted H_2/H_∞ norms.

In this paper we present an overview of the application of statistical learning theory in a robust control framework. The paper is organized as follows: in Section II some intractable problems that arises in control theory are reviewed. Section III contains the probabilistic robust synthesis problem formulation. In Section IV different notions of approximation of function minima is given. In Section IV we investigate the role of empirical process theory in designing robust controllers. Sections VI and VII contain algorithm based on different notions of near minimum or maximum. Finally, in subsequent section we conclude and discuss the scope of research.

II. NP HARD PROBLEMS IN CONTROLLER SYNTHESIS

During the past decade researchers studied the computational complexity of controller analysis and synthesis problems. Surprisingly these studies showed that a number of analysis and synthesis problems are NP-hard problems. Checking robust stability, robust positive semi-definiteness, robust norm boundedness, robust non-singularity; all these problems were proved as NP Hard. These are all analysis problems and intuitively if the analysis problem is NP hard then the corresponding synthesis problem is at least as difficult. In the following we give some of the NP hard synthesis problems.

A. Robust Stabilization Against Structured Perturbation

Consider the problem of finding a fixed controller K which stabilizes an uncertain plant P, subjected to structured perturbations around P_0. The well known solution to this problem is to validate if an associated structured singular value is less than one. It is shown in [1] that answering this decision problem is NP hard. It is also shown in [2] that even to find an approximate solution is also NP Hard.

B. Constant Output Feedback Stabilization with Constraints

For a given state space system matrices A, B, C it is NP hard problem to decide whether there exists a controller matrix K whose elements are bounded within an interval,

This work is supported by UKIERI project

Titas Bera is a research scholar in the Department of Aerospace Engineering, Indian Institute of Science, Bangalore - 560012, India (e-mail: titasbera@aero.iisc.ernet.in)

M. S. Bhat and D. Ghose are with the Department of Aerospace Engineering, Indian Institute of Science, Bangalore - 560012, India (e-mail: msbdcl@aero.iisc.ernet.in;dghose@aero.iisc.ernet.in)

stabilizes the closed loop system or not [1]. One can try to solve the problem using Tarski-Seidenberg elimination theory but the number of polynomial inequalities becomes exponential with the dimension of the system.

C. Simultaneous Stabilization using Constant Output Feedback

For a family of plants (a family of A,B,C matrices) it is NP hard to problem to decide whether there exists a controller matrix K whose elements are bounded within an interval, stabilizes the closed loop system or not [1].

These results leads to a realization that such innocent looking robust synthesis problems are in fact intractable. To tackle the situation randomized algorithm proves viable. The empirical process theory(or its widespread name statistical learning theory) gives us justification to use such algorithms.

III. Probabilistic Robust Synthesis Problem

Let there exists a family of plants defined as $\{G(x), x \in X\}$ parameterized by x. Also assume that there exists a controller family $\{K(y), y \in Y\}$ parameterized by y. Now, suppose P is a probability measure on the set X. Let, $\varepsilon \in \mathbb{R}$ be a given accuracy parameter.

In the probabilistic framework, instead of looking for a $K(y)$ which satisfies worst case performance for all the plant instances, one compromises the search for a $K(y)$ that satisfies the performance index for most of the plant instances except possibly for those belonging to a set of measure no larger than ε. The performance function to be minimized can be defined as

$$J := E[\psi(G(x), K(y))] \tag{1}$$

where, $E(\cdot)$ denotes the expected value. The expectation measure with respect to P of cost functional captures the intuitive idea that a controller is allowed to perform poorly for some instances. Notice that the cost function becomes a function of controller parameter once the underlying probability distribution P of plant family is chosen, i.e.

$$f(y) := E_P[\psi(G(x), K(y))] \tag{2}$$

Let us define $g_y(x) := \psi(G(x), K(y))$. Also, without loss of generality for each $y \in Y$, $g_y(\cdot)$ maps X into $[0,1]$. Let, $\Gamma := \{g_y(\cdot), y \in Y\}$ is the associated family of controllers. The objective is to find $y = y^*$ that minimizes $f(y)$.

$$y^* = \arg \min_{\substack{y \\ x \in X \\ g_y \in \Gamma}} E_P[g_y(x)] \tag{3}$$

IV. Notions of Near minima

In the previous section, an abstract formulation of robust controller synthesis problem was given, which ultimately lead to the problem of finding the minimum of some function $f : Y \to [0,1]$. However, since the problem of finding the absolute minimum is NP hard one needs to find a near minimum of $f(\cdot)$. In the following, we describe different notions of minima as mentioned in [9].

A. True Minimum

Definition: Suppose $f : Y \to \mathbb{R}$. Then the true minimum of $f(\cdot)$ is defined as,

$$f^* := \min_{y \in Y} f(y) \tag{4}$$

B. Approximate Near Minimum

Definition: Suppose $f : Y \to \mathbb{R}$ and that $\varepsilon > 0$ is a given number. A number $f_0 \in \mathbb{R}$ is said to be an approximate near minimum of $f(\cdot)$ to the accuracy of ε if,

$$\inf_{y \in Y} f(y) - \varepsilon \leq f_0 \leq \inf_{y \in Y} f(y) + \varepsilon \tag{5}$$

This is the most intuitive notion of near minima. However, as proved in [16], it is even NP hard to find a near approximation of f^*. This means with any given number ε it is NP hard to decide whether $|f_0 - \inf_{y \in Y} f(y)| \leq \varepsilon$ or not. This necessitates to look for other notions of near minima.

C. Probable Near Minimum

Definition: Suppose $f : Y \to \mathbb{R}$, P is a given probability measure on the set X and $\alpha > 0$ is a given number. A number f_0 is said to be a probable near minimum of $f(\cdot)$ to level α if $f_0 \geq f^*$ and in addition if $P\{y \in Y : f(y) < f_0\} \leq \alpha$. Since, it is NP hard to decide *exactly* whether a given number f_0 is ε distance away from the true minimum or not, therefore most intuitive choice is to assign a certain degree of uncertainty to f_0 to be near minimum. The interpretation is, there could be an exceptional set S whose measure is at most α such that

$$\inf_{y \in Y} f(y) \leq f_0 \leq \inf_{y \in Y \setminus S} f(y) \tag{6}$$

D. Probable Approximate Near Minimum

Definition: Suppose $f : Y \to \mathbb{R}$, that P_Y is a given probability measure on Y and that $\varepsilon, \alpha > 0$ are given numbers. A number $f_0 \in \mathbb{R}$ is said to be a near minimum of $f(.)$ to accuracy ε and level α, or a probably approximate minimum of $f(.)$ to accuracy ε and level α if $f_0 \geq f^* - \varepsilon$ and in addition $P_Y\{y \in Y : f(y) < f_0 - \varepsilon\} \leq \alpha$.

Probable near minimum and probable approximate near minimum are compromises from evaluating true near minimum or ε-accurate true near minimum. But before evaluating these one needs to calculate the expectation of cost functional for a fixed controller parameter vector and plant probability distribution P. This requires computation of multidimensional volume integral which is almost impossible to compute except in some trivial cases. Also, any deterministic numerical computation technique for calculating such integrals has exponential complexity with the number of dimension. Therefore to calculate expectation one has to rely on randomized algorithms which have strong mathematical support arising from empirical process theory.

Different notions of approximate near minimum leads to different randomized algorithms. In the following we describe briefly different algorithms, details of which can be found in [9] and [7].

V. Role of Empirical Process in Approximate Minima Calculation

Empirical process theory tells us that we can approximate a quantity with arbitrarily small error from empirical experiments/observations. To understand how this can help in a way to obtain an unbiased estimate of true minimum and consequently to find lower bounds on sample complexities, let us consider the case of a binary classification problem where some training examples $(x_1,y_1),...(x_m,y_m)$ were given and one needs to construct the relationship between the training samples (predicting the underlying function $h(x)=y$) and to predict the outcome for a new sample x. One can define an empirical risk function $\zeta_i := \frac{1}{2}|h(x_i)-y_i|$ which is either 0 or 1 (provided we have a ± 1 valued function h). All training samples are drawn independently, so we are faced with independent Bernoulli trials. The $\zeta_1,...\zeta_m$ are i.i.d random variable. A famous inequality due to Chernoff characterizes how the empirical mean of loss function defined as $\frac{1}{m}\sum_{i=1}^{m}\zeta_i$ converges to the expected value of loss function ζ, denoted as $E(\zeta)$ for a particular h,

$$P\{|\frac{1}{m}\sum_{1}^{m}\zeta_i - E(\zeta)| \geq \varepsilon\} \leq 2exp(-2m\varepsilon^2) \tag{7}$$

This indicates that the convergence in probability to the true mean is exponentially fast as the number of observations increases. This is also known as law of large numbers. A similar inequality called Hoeffding inequality also exists and is stated as follows.

Theorem 1: Hoeffding [11] Let $\zeta_i, i \in [m]$ be m independent instances of a bounded random variable ζ, with values in $[a,b]$. Let their average be $Q_m = \frac{1}{m}\sum_{i=1}^{m}\zeta_i$. Then for any $\varepsilon > 0$,

$$P\{|Q_m - E(\zeta)| \geq \varepsilon\} \leq \exp(\frac{-2m\varepsilon^2}{(b-a)^2}) \tag{8}$$

From the Chernoff bound one can find how many observations are required to estimate the unknown quantity $E(\zeta)$ to an accuracy of ε and confidence of $1-\delta$, which is

$$m \geq \frac{1}{2\varepsilon^2}\ln\frac{2}{\delta} \tag{9}$$

Ideally, we want to compute $f(y)$ for a given y so that $f(\cdot)$ can be minimized using any suitable method, but except in some trivial situations it is rather difficult to compute $f(y)$ exactly. Instead, we can approximate $f(y)$ empirically with an arbitrary small error. One can have a collection of m i.i.d samples in X, generated according to (11) and can define for each function $g_y \in Y$, the empirical mean based on a multi-sample x as

$$\hat{f}(y) := \frac{1}{m}\sum_{j=1}^{m} g_y(x_j) \tag{10}$$

In other words, the actual performance $f(y)$ of a controller $K(y)$ is approximated by its average performance on the randomly generated m plants. One can further try to calculate $\inf_{y\in Y}\hat{f}(y)$ but there exists two crucial caveat.

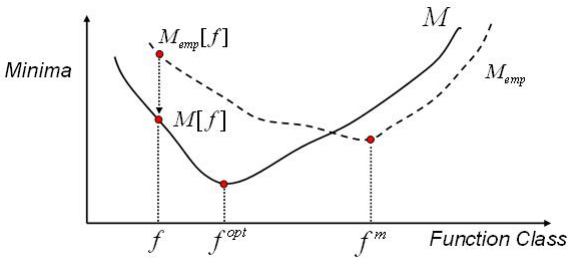

Fig. 1. The convergence of empirical minima to actual minima. The horizontal axis gives a one dimensional representation of the function class; the vertical axis denotes the minima. For each *fixed* function f, the law of large numbers tells us that as the sample size goes to infinity, the empirical minima $M_{emp}[f]$ converges toward the true minima $M[f]$ (indicated by the downward arrow). This does not imply, however, that in the limit of infinite sample sizes, the minimizer of the empirical minima, f^m, will lead to a value that is as good as the best attainable minima, $M[f^{opt}]$(consistency). For the latter to be true, we require convergence of $M_{emp}[f]$ towards $M[f]$ to be uniform over all functions.

First, the important property of Chernoff bound (or more general Hoeffding bound) is that they are probabilistic in nature. They do not rule out the possibility of existence of some functions for which the error convergence inequality does not hold. To tackle the situation for such erratically behaving function we need to have a more *consistent* law of large numbers. This means that to determine the consistency of empirical risk minimization one need to consider the worst case over all functions, see fig 1. In other words, the law of large numbers has to be uniform over all such functions. This insight is brought by Vapnik and Chervonenkis in their seminal work [12] and the following theorem describes the same.

Theorem 2: Vapnik-Chervonenkis One sided uniform convergence in probability

$$\lim_{m\to\infty} P\{\sup_{g_y\in\Gamma}|\hat{f}(y) - f(y)| > \varepsilon\} = 0 \tag{11}$$

for all $\varepsilon > 0$, is a necessary and sufficient condition for non trivial consistency of empirical risk minimization. The condition of uniform convergence depends on the set of functions for which it must hold.

Secondly, one cannot possibly try to evaluate the empirical expectation for each possible $f(y)$ from the infinite function class Γ.

VI. Randomized Algorithm for Finding Probably Approximate Near Minima

Now for the first problem, let us define

$$q(m,\varepsilon,\Gamma) := P^m\{x \in X^m : \sup_{g_y\in\Gamma}|\hat{E}(g_y;x) - E_p(g_y)| > \varepsilon\} \tag{12}$$

If the family of functions show the property of $q(m,\varepsilon;\Gamma)\to 0$ as $m\to\infty$ for each $\varepsilon > 0$ then the family has the uniform convergence of empirical mean (UCEM) property. Depending on whether the function class possess the UCEM property,

different randomized algorithms to approximate probable near minima can be obtained. Assume that the function class G posses uniform convergence of empirical means property. If such is the case then one can choose m large enough according to (11) such that it can be said with confidence $1-\delta$ that

$$|f(y)-\hat{E}(g_y;x)| \leq \varepsilon, \forall y \in Y \tag{13}$$

In other words, the function $\hat{E}(g;x)$ is a uniformly close approximation to the original objective function $f(\cdot)$. Hence it really follows that an exact minimizer of $\hat{E}(g;x)$ is also an approximate near minimizer of $f(\cdot)$ to accuracy ε.

For the second problem, since there might not be a closed-form expression for $f(y)$, therefore efficient gradient based minimization methods turns out to be impossible to use. One obvious way out is to limit the infinite function class into a finite one by selecting n number of i.i.d samples of controller g_y form Γ according to a distribution Q and therefore calculating the empirical minimum of the expected value of $f(y)$ over all $g_y \in \Gamma$. Since, we are not assuming the consistency of the function class Γ therefore this empirical minimum could be far away from the true minimum of $f(y)$. Specifically, let g_y^* denote the controller from the function class Γ that gives the lowest value $\hat{f}(y)$. Define $c^* := \hat{f}(y)$. Also, define $M(c^*) := \{g_y \in \Gamma : \hat{f}(y) < c^*\}$ i.e. $M(c^*)$ is the set of functions for which $\hat{f}(y) < c^*$. Then given a positive constant α, it can be said with confidence $(1-\alpha)^n$ that the measure $Q(M(c^*) \leq \alpha$. In other words, it can be said with confidence $(1-\alpha)^n$ that the controller g_y^* minimizes the performance function over *nearly all* of Γ.

Notice that the calculation of confidence interval of the approximation of true minima involves double randomization of both the controller and plant family. Therefore, let us partition the confidence interval of $1-\delta$ into two halves and estimate the corresponding sample complexities.

Now, according to Hoeffding inequality,

$$P^m\{x \in X^m : |\hat{E}(f;x) - E_P(f)| > \varepsilon\} \leq 2\exp(-2m\varepsilon^2) \tag{14}$$

where, P^m is the m^{th} fold product of the probability measure. Now let us find out the *union bound* for those set of samples in X for which the true minimum and empirical minimum could differ by more than ε for all n number of functions. Therefore, let

$$C_\varepsilon^i := \{x \in X^m : |\hat{E}(f_i;x) - E_P(f_i)| > \varepsilon\} \tag{15}$$

denote the set of samples for which the true minimum differs from the actual one for a particular f_i. Then

$$P(C) = P(C_\varepsilon^1 \cup \ldots \cup C_\varepsilon^n) \leq \sum_{i=1}^{n} P(C_\varepsilon^i) \tag{16}$$

Then, from the Hoeffding inequality

$$P(C) \leq 2n\exp(-2m\varepsilon^2) \tag{17}$$

Therefore, from the following inequalities

$$(1-\alpha)^n \leq \delta/2 \tag{18}$$

$$2n\exp(-2m\varepsilon^2) \leq \delta/2 \tag{19}$$

one can derive the sample complexity as,

$$n \geq \frac{\ln(2/\delta)}{\ln[1/(1-\delta)]} \tag{20}$$

$$m \geq \frac{1}{2\varepsilon^2}\ln\frac{4n}{\delta} \tag{21}$$

Next we describe the following algorithm based on above arguments.

Algorithm 1:
Choose integers $n \geq \frac{\ln(2/\delta)}{\ln[1/(1-\delta)]}$ and $m \geq \frac{1}{2\varepsilon^2}\ln\frac{4n}{\delta}$
Generate n i.i.d samples $x_1, x_2, \ldots x_N$ from the set X according to (P) and m i.i.d samples $y_1, y_2, \ldots y_m$ from Y according to P_k

For $j \leftarrow 1$ to M do
calculate $\hat{f}_i = \frac{1}{m}\sum_{i=1}^{m} \psi(k(y_j), G(x_i))$
Return The empirical optimum controller
$\hat{K}(y_j) = \arg\min_i \hat{f}_i$
$f_0 := \min_i \hat{f}_i$

Then, with confidence $1-\delta$, it can be said that $\hat{f}_0$ is a probably approximate near minimum of $f(.)$ to accuracy ε and level α.

Note that in this algorithm the associated function class Γ is not assumed to have any property or structure. It is rather a general algorithm. The caveat is that the empirical minimization of cost functional heavily depends on the sampled controller sequence. Also, as the level parameter α approaches zero, both the n and m parameter has to be increased simultaneously. The performance of the controller depends on the samples directly.

Now consider the case where the function class Γ has the consistency property. We need to find a bound of the following probability $P\{\sup_{g_y \in \Gamma}(f(y) - \hat{f}(y)) > \varepsilon\}$. Now again consider the case of binary classification. Each function of the class separates the training samples in a certain way and thus induces a certain labelling of the samples. Since the labels are in $\{\pm 1\}$, there are at most 2^m different labelling for m samples. A very rich function class might be able to realize all 2^m separations, in which case it is said to *shatter* the m samples. However the given class of functions might not be sufficiently rich to shatter the m points. The Vapnik- Chervonenkis dimension or VC-dimension is defined as the largest m such that there exists a set of m points which the class can shatter. A necessary and sufficient condition of a collection of sets $\{g_y \in Y\}$ to have the uniform convergence of empirical mean property under *any* probability distribution P, is that the corresponding set's VC dimension should be finite. This means that no matter how we choose the m controllers, the empirical mean will converges towards true mean when number of samples tends to infinity. The convergence property now does not depend on the sampled controller sequences. Based on this we can construct another randomized algorithm. We begin by partitioning the confidence δ into two parts as in previous case. Since the collection of sets $\{g_y \in \Gamma\}$ has the uniform

convergence of empirical mean property therefore it has a finite VC dimension d. Then as proved in [9] if we draw at least

$$n \geq \max\{\frac{16}{\varepsilon^2}\ln\frac{8}{\delta}, \frac{32d}{\varepsilon^2}\ln\frac{32e}{\varepsilon^2}\} \tag{22}$$

i.i.d samples of plants then we can say with confidence $1-\delta/2$ that each empirical estimate $\hat{f}(y)$ is within ε of the corresponding true value $f^*(y)$. Next let us choose an integer m large enough that $(1-\alpha)^m \leq \delta/2$, or

$$m \geq \frac{\ln(2/\delta)}{\ln(1/(1-\alpha))} \tag{23}$$

Choose i.i.d samples $g_{y_1}, \ldots g_{y_m}$ distributed according to any Q and pick g_{y_j} that gives the minima for $\hat{f}(y)$. This empirical minimum, is within confidence $1-\delta/2$, a near minimum of $\hat{f}(y)$. Combining both statements shows that this procedure gives a near minimum of $f(y)$ to accuracy ε, confidence $1-\delta$, and to a level α. In the following we present the algorithm based on above arguments.

Algorithm 2:
Choose integers $m \geq \frac{\ln(2/\delta)}{\ln[1/(1-\delta)]}$ and $n \geq \max\{\frac{16}{\varepsilon^2}\ln\frac{8}{\delta}, \frac{32d}{\varepsilon^2}\ln\frac{32e}{\varepsilon^2}\}$
Generate n i.i.d samples $x_1, x_2, \ldots x_n$ from set X according to (P) and m i.i.d samples $y_1, y_2, \ldots y_m$ from Y according to P_k
For $j \leftarrow 1$ to M do
calculate $\hat{f}_i = \frac{1}{M}\sum_{i=1}^{M}\psi(K(y_j), G(x_i))$
Return The empirical optimum controller
$\hat{K}(y_j) = \arg\min_j \hat{f}_i$

There can be two possible source of conservatism in the estimates of sample complexity, either in the estimate of VC dimension d or the estimate of sample complexity for the VC dimension d. Therefore, there is considerable room for improvement in the estimate as shown in [13].

VII. Randomized Algorithm for Finding Probable Near Maxima

Efficient algorithms based on the concept of probable near minimum is introduced by Tempo et al, [15] and Khargonekar and Tikku [14]. Here we briefly describe their approach.

Define the distribution function of the random variable f as P. Now, for each $a \in \mathbb{R}$ let,

$$r(a) := P\{y \in X : f(y) \leq a\} \tag{24}$$

also,

$$\bar{f}(y) := \max_{1\leq i\leq m} f(y_i) \tag{25}$$

Given $\varepsilon > 0$ define $a_\varepsilon := \inf\{a : r(a) \geq 1-\varepsilon\}$. This means $r(a) < 1-\varepsilon$ if $a < a_\varepsilon$. Now suppose $\bar{f}(y) \geq a_\varepsilon$ then

$$P\{y \in Y : f(y) \leq \bar{f}(y)\} = r[\bar{f}(y)] \geq 1-\varepsilon \tag{26}$$

or,

$$P\{y \in Y : f(y) > \bar{f}(y)\} = 1 - r[\bar{f}(y)] \leq \varepsilon \tag{27}$$

Therefore, if $P\{y \in Y : f(y) > \bar{f}(y)\} = 1 - r[\bar{f}(y)] > \varepsilon$ then $\bar{f}(y) < a_\varepsilon$ and $r[\bar{f}(y)] < 1-\varepsilon$. Since, $\bar{f}(y) < a_\varepsilon$, therefore $f(y_i) < a_\varepsilon$. Taking the *mth* fold probability is simply

$$P^m\{y \in Y : P\{f(y) > \bar{f}(y)\} > \varepsilon\} \leq (1-\varepsilon)^m \tag{28}$$

Now choosing an integer m such that $(1-\varepsilon)^m \leq \delta$ leads to a lower bound of m as

$$m \geq \frac{\ln(1/\delta)}{\ln[1/(1-\varepsilon)]} \tag{29}$$

Therefore if this is the lower bound of the sample complexity then it can be said with confidence at least $1-\delta$ that $\bar{f}(y)$ is a probable near minimum of $f(\cdot)$ to level α. Notice that the bound in sample complexity is independent of number of uncertain parameters and the density function of $f(y)$. This is best possible estimate for m. Here we present the following algorithm based on the above argument.

Algorithm 3:
Choose integers $m \geq \frac{\ln(1/\delta)}{\ln[1/(1-\varepsilon)]}$
Generate m i.i.d samples $y_1, y_2, \ldots y_m$ from Y according to P
For $j \leftarrow 1$ to m do
calculate $\hat{f}_i = \min_{1\leq i\leq m}\psi(K(y_i), G(x_i))$
Return The empirical optimum controller
$\hat{K}(y_i) = \arg_i \hat{f}_i$

Notice that in this algorithm only controller randomization is used. In this sense the algorithm computes probabilistically *worst* case performance bounds instead of an average case bound. As one can expect this reduces the sample complexity greatly as reported in [15].

VIII. Conclusions and Future Work

In this paper we reviewed several existing empirical process theory based randomized algorithm for probabilistic robust controller synthesis problem. The sample complexity of algorithm 1 and 3 are found best possible, while there exists considerable scope to improve sample complexity bounds for algorithm 2.

References

[1] V. D. Blondel and J. N. Tsitsiklis, A survey of computaional complexity results in systems and control *Automatica*, vol. 36, no. 9, 2000, pp 1249-1274.
[2] V. D. Blondel and J. N. Tsitsiklis, NP-hardness of some linear control design problems *SIAM Journal on Control and Optimization*, vol. 35, 1997, pp 2118-2127.
[3] D. S. Bernstein, Some open problems in matrix theory arising in linear systems and control, *Linear Algebra and Applications*, vol. 162, 1992, pp 409-432
[4] A. Nemirovskii, Several NP-hard problems arising in robust stability analysis, *Mathematics of Control Signal and Systems*, vol. 6, no. 1, 1993, pp 99-105
[5] S. Poljak, J. Rohn, Checking robust nonsingularity is NP-hard, *Mathematics of Control Signal and Systems*, vol. 6, no. 1, 1991, pp 1-9
[6] M. R. Garey and D. S. Johnson, *Computers and Intractability: A guide to the theory of NP-completeness*, Freeman, New York, 1979.
[7] R. E. Tempo, G. Calafiore and F. Dabbene, *Randomized ALgorithms for Analysis and Control of Uncertain Systems*, Springer, 2005.
[8] M. Vidyasagar, Statistical learning theory and randomized algorithms for control *IEEE Control Systems Magazine*, vol. 18, no. 6, 1998, pp 69-85.

[9] M. Vidyasagar, *Learning and Generalisation: with application to neural network*, Springer-Verlag, 2003.
[10] M. Vidyasagar and V. D. Blondel, Probabilistic solutions to some NP-hard matrix problems, *Automatica*, vol. 37, 2001, pp 1397-1405
[11] W. Hoeffding, Probability inequalities for sums of bounded random variable, *Journal of American Statistical Association*, vol. 58, 1963, pp 13-30
[12] V. N. Vapnik, Y. A. Chervonenkis, Necessary and sufficient conditions for the uniform convergence of means to their expectations, *Theory of Probability and its Applications*, vol. 26, no. 3, 1981, pp 532-553
[13] A. Kowalczyk, H.Ferra, J. Szymanski, Combining statistical physics with VC bounds on generalisation in learning systems, *Proceedings of 6th Australian Conference on Neural Networks, Sydney*, 1995, pp 41-44
[14] P. P. Khargonekar and A. Tikku, Randomized Algorithm for robust control analysis have polynomial time complexity *in Proceedings of the* 35^{th} *Conference on Decision and Control*, Kobe, Japan, December 1996.
[15] R. E. Tempo, W. Bai and F. Dabbene, Probabilistic robustness analysis: Explicit bounds for the minimum number of sampling points*System and Control Letters*, vol. 30, 1997, pp 237-242.
[16] J. N. Tsitsiklis, Complexity theoritic aspects of problems in control theory, *Technical Report LIDS - P- 2203*, Laboratory for information and decision systems, M.I.T, Cambridge, MA, 1993